



# STIC Search Report

## Biotech-Chem Library

File Copy  
09/974,760  
updated

STIC Database Tracking Number: 145470

TO: David Lamberston  
Location: rem/2b79/2c70  
Art Unit: 1636  
Thursday, February 17, 2005

Case Serial Number: 09/974760

From: Barb O'Bryen  
Location: Biotech-Chem Library  
Remsen 1A69  
Phone: 571-272-2518 *BOB*

barbara.obryen@uspto.gov

### Search Notes

result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2035	82.9	469	6	ABR64151	ABr64151 Aspergill	
2	830	33.8	459	5	AAU10871	Aau10871 Penicilli	
3	830	33.8	459	7	ADe881191	ADe881191 Mlcrp SEQ	
4	157.5	6.4	424	5	ABP63135	ABp63135 lovP gene	
5	157.5	6.4	424	5	ABP35719	ABp35719 Fungal ZB	
6	154.5	6.3	919	5	ABP35687	ABp35687 Fungal ZB	
7	154.5	6.3	919	8	ADN18782	ADn18782 Bacterial	
8	141	5.7	727	7	ADC52108	ADc52108 A. oryzae	
9	136	5.5	604	2	AAW37163	AAw37163 Aspergill	
10	136	5.5	604	5	ABP35591	ABp35591 Fungal ZB	
11	135	5.5	433	5	ABP35583	ABp35583 Fungal ZB	
12	132	5.4	1170	5	ABP35667	ABp35667 Fungal ZB	
13	132	5.4	1170	8	ADS43829	ADs43829 Bacterial	
14	131.5	5.4	480	5	ABP35612	ABp35612 Fungal ZB	
15	130.5	5.3	662	5	ABP35590	ABp35590 Fungal ZB	
16	130.5	5.3	792	5	ABG93477	ABg93477 Aspergill	
17	129.5	5.3	579	5	ABP35589	ABp35589 Fungal ZB	
18	126.5	5.2	341	2	AAW58573	AAw58573 Trichoder	
19	126	5.1	821	5	ABG93475	ABg93475 Aspergill	
20	124.5	5.1	736	5	ABP35614	ABp35614 Fungal ZB	
21	124	5.0	984	5	ABP35627	ABp35627 Fungal ZB	
22	123.5	5.0	437	5	ABP35582	ABp35582 Fungal ZB	
23	123.5	5.0	607	5	ABP35705	ABp35705 Fungal ZB	
24	122	5.0	821	5	ABG93469	ABg93469 Aspergill	
25	122	5.0	821	5	ABP35586	ABp35586 Fungal ZB	

FI  
FT

Aab300811	Amino acid
Aab35696	Fungal ZB2
Adk26832	Disease c
Aap33710	Fungal ZB2
Aap33575	Fungal ZB2
Adg93472	Aspergill
Adg93472	Aspergill
Adg84193	Orf1, SQS
Aap336476	Fungal ZB2
Aap336468	Fungal ZB2
Abt84508	Human ani
Abt84509	Mouse ani
Abt84509	Mouse ani
Aap335652	Fungal ZB2
Aap335652	Fungal ZB2
Adg71247	Novel hbm
Abt84465	Human can
Abt84465	Human can
Adg64588	Novel hbm
Adg64588	Novel hbm
Adt10449	Human pro
Aap336508	Fungal ZB2
Aae316891	Plecteur

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2005, 01:57:35 ; Search time 43 Seconds  
(without alignments)  
814.196 Million cell updates/sec

Title: US-09-974-760B-91

Perfect score: 2456

Sequence: 1 MAADQIGFTSVTLSPVEGS.....FSLARKKHGMLRLNNIPP 469

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 95%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgn2\_6/ptodata/1/iaa/5A COMB.pcp.\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B COMB.pcp.\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A COMB.pcp.\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B COMB.pcp.\*
  - 5: /cgn2\_6/ptodata/1/iaa/PTUS.COMB.pcp.\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	157.5	6.4	424	US-10-029-180-104	Sequence 104, App
2	136	5.5	604	US-09-197-814-3	Sequence 3, Appli
3	136	5.5	604	US-09-920-581-3	Sequence 3, Appli
4	119	4.8	631	US-09-147-119-7	Sequence 7, Appli
5	119	4.8	707	US-09-949-016-8495	Sequence 8495, Ap
6	118	4.8	439	US-09-248-796A-18964	Sequence 18964, A
7	116	4.7	989	US-09-248-796A-19109	Sequence 19109, A
8	112.5	4.6	685	US-09-248-796A-20197	Sequence 20197, A
9	111.5	4.5	986	US-09-248-796A-19088	Sequence 19088, A
10	109	4.4	534	US-09-248-796A-19568	Sequence 19568, A
11	108.5	4.4	314	US-09-248-796A-18676	Sequence 18676, A
12	108	4.4	835	US-09-252-991A-32121	Sequence 32121, A
13	107	4.4	479	US-09-248-796A-17560	Sequence 17560, A
14	106.5	4.3	775	US-08-714-070A-1	Sequence 1, Appli
15	106	4.3	971	US-09-248-796A-19531	Sequence 19531, A
16	104	4.2	335	US-09-248-796A-17099	Sequence 17099, A
17	104	4.2	1341	US-09-949-016-6890	Sequence 6890, Ap
18	104	4.2	1344	US-09-949-016-10925	Sequence 10925, A
19	102.5	4.2	808	US-09-248-796A-18679	Sequence 18679, A
20	101.5	4.1	351	US-09-248-796A-20887	Sequence 20887, A
21	101.5	4.1	421	US-09-248-796A-19591	Sequence 19591, A
22	101	4.1	515	US-09-252-991A-22451	Sequence 22451, A
23	101	4.1	1587	US-09-845-583A-10	Sequence 10, Appli
24	101	4.1	1587	US-09-561-709B-3	Sequence 3, Appli
25	100	4.1	758	US-10-029-180-92	Sequence 92, Appli
26	99	4.0	1065	US-09-538-092-84	Sequence 84, Appli
27	98.5	4.0	309	US-10-029-180-98	Sequence 98, Appli

ALIGNMENTS

RESULT 1

US-10-029-180-104  
; Sequence 104, Application US/10029180  
; Patent No. 6806082

GENERAL INFORMATION:

- APPLICANT: Cali, Brian M.
- APPLICANT: Holtzman, Doug
- APPLICANT: Madden, Kevin T.
- APPLICANT: Milna, G. Todd
- APPLICANT: Sherman, Amir
- APPLICANT: Silva, Jeffry C.
- APPLICANT: Trueheart, Josh
- APPLICANT: Zhang, Lixin
- TITLE OF INVENTION: No. 6806082el Regulators of Fungal Gene Expression
- FILE REFERENCE: MIC-004
- CURRENT APPLICATION NUMBER: US/10/029,180
- CURRENT FILING DATE: 2001-12-22
- PRIOR FILING DATE: 2001-12-22
- PRIOR FILING DATE: 2000-12-22
- NUMBER OF SEQ ID NOS: 138
- SOFTWARE: FastSeq for Windows Version 4.0
- SEQ ID NO 104
- LENGTH: 424
- TYPE: PRT
- ORGANISM: Artificial Sequence
- FEATURE:
- OTHER INFORMATION: fungal gene

US-10-029-180-104

Query Match 6.4%; Score 157.5; DB 4; Length 424;

Best Local Similarity 20.6%; Pred. No. 2.1e-07;  
Matches 94; Conservative 62; Mismatches 161; Indels 139; Gaps 22;

QY	32	RRSCDRCHAQKIKCTGNKEVTGRAPCQRCQQAAGLCVTVSERCPCKKLRQRA-----	83
DB	7	RIACDRCHGQKLRCH-----SGGGFCVRCACAKATCSWSQSLSRNLKHNAPISDVPLA	62
QY	84	ADLYSADPDP-----CLHMSGP-----PVPSQSLPLDVS-ES	114
DB	63	CAQLATQSDPTNPQFGAYMSQPSAGVDIDINLLQTFDTSTPWALPAGRYVPSAQSM	122
QY	115	HSSNTRQFLDPPDSYDWSWTSIGT-----DEAIDTDCWGLSCDGGFSCQLEP	163
DB	123	ETVNYGHTHEADLPATADWMPAVANGPVQTTTPANWQQAFOE-WAM-----MASQHPV	175
QY	164	TEPLDLPSPFESTVEKAPLPVSSDIARAASAQRELFDLSA-----VSQLEETILLAV	216
DB	176	ATMDTFSR-TSPVSDAVDPKTKVCLLATIRELSNVDLYAHEATVPKPPASLEB-----	229
QY	217	TVEWPKQEIWTHPIGMFFNARRLLTVLRQQ-----AQADCHQGT-----	256

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2005, 02:03:56 ; Search time 133 Seconds

(without alignment)

1152.221 Million cell updates/sec

Title: US-09-974-760B-91

Perfect score: 2456

Sequence: 1 MAADQGIFTNSVTLSPVSGS.....FSLARKKHGMLRLNNIPP 469

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1376875 seqs, 326749119 residues

Total number of hits satisfying chosen parameters: 1376809

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 95%

Listing first 45 summaries

Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1858.5	75.7	366	16	US-10-402-056-105
2	830	33.8	459	10	US-09-836-705-42
3	157.5	6.4	424	13	US-10-029-180-104
4	157.5	6.4	424	15	US-10-149-310-302
5	154.5	6.3	919	15	US-10-369-493-1435
6	154.5	6.3	919	15	US-10-149-310-238
7	136	5.5	604	9	US-09-920-581-3
8	136	5.5	604	14	US-10-371-421-3
9	136	5.5	604	15	US-10-149-310-46
10	135	5.5	433	9	US-09-801-368-38
11	135	5.5	433	15	US-10-149-310-30
12	132	5.4	1170	15	US-10-369-493-22259
13	132	5.4	1170	15	US-10-149-310-198

14	131.5	5.4	480	15	US-10-149-310-88	Sequence 88, Appl
15	130.5	5.3	662	9	US-09-801-368-42	Sequence 42, Appl
16	130.5	5.3	662	15	US-10-149-310-44	Sequence 44, Appl
17	130.5	5.3	792	16	US-10-468-250A-130	Sequence 130, Appl
18	129.5	5.3	579	15	US-10-149-310-42	Sequence 42, Appl
19	126	5.1	821	16	US-10-468-250A-128	Sequence 128, Appl
20	124.5	5.1	736	15	US-10-149-310-92	Sequence 92, Appl
21	124	5.0	984	15	US-10-149-310-118	Sequence 118, Appl
22	123.5	5.0	437	15	US-10-149-310-28	Sequence 28, Appl
23	123.5	5.0	607	15	US-10-149-310-274	Sequence 274, Appl
24	122	5.0	821	15	US-10-149-310-36	Sequence 36, Appl
25	122	5.0	821	16	US-10-468-250A-122	Sequence 122, Appl
26	121.5	4.9	964	15	US-10-149-310-256	Sequence 256, Appl
27	121	4.9	470	15	US-10-149-310-284	Sequence 284, Appl
28	121	4.9	739	9	US-09-801-368-396	Sequence 396, Appl
29	121	4.9	739	15	US-10-149-310-214	Sequence 214, Appl
30	121	4.9	821	16	US-10-468-250A-125	Sequence 125, Appl
31	120.5	4.9	795	16	US-10-468-250A-129	Sequence 129, Appl
32	120	4.9	470	15	US-10-149-310-160	Sequence 160, Appl
33	119	4.8	470	15	US-10-149-310-168	Sequence 168, Appl
34	119	4.8	565	15	US-10-149-310-188	Sequence 188, Appl
35	118.5	4.8	857	15	US-10-149-310-80	Sequence 80, Appl
36	118.5	4.8	2016	17	US-10-488-056-52	Sequence 52, Appl
37	117	4.8	473	15	US-10-149-310-170	Sequence 170, Appl
38	117	4.8	767	15	US-10-149-310-90	Sequence 90, Appl
39	116.5	4.7	377	9	US-09-853-386-74	Sequence 74, Appl
40	116.5	4.7	444	15	US-10-149-310-34	Sequence 34, Appl
41	116.5	4.7	882	15	US-10-149-310-122	Sequence 122, Appl
42	116.5	4.7	1445	15	US-10-149-310-110	Sequence 110, Appl
43	116	4.7	470	15	US-10-149-310-164	Sequence 164, Appl
44	115.5	4.7	384	15	US-10-149-310-32	Sequence 32, Appl
45	115	4.7	517	9	US-09-801-368-100	Sequence 100, Appl

#### ALIGNMENTS

RESULT 1  
US-10-402-056-105  
; Sequence 105, Application US/10402056  
; Publication No. US20040191877A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, Amir  
; APPLICANT: Roberts, Shannon  
; APPLICANT: Trueheart, Joshua  
; APPLICANT: Milne, G. Todd  
; TITLE OF INVENTION: LOVE VARIANT REGULATOR MOLECULES  
; FILE REFERENCE: 14184-029001  
; CURRENT APPLICATION NUMBER: US/10/402,056  
; CURRENT FILING DATE: 2003-03-28  
; PRIOR APPLICATION NUMBER: PCT/US02/32248  
; PRIOR FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US 09/974,760  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US 60/328,339  
; PRIOR FILING DATE: 2001-10-09  
; NUMBER OF SEQ ID NOS: 118  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 105  
; LENGTH: 366  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: synthetically generated variant  
US-10-402-056-105

Query Match 75.7%; Score 1858.5; DB 16; Length 366;

Best Local Similarity 77.6%; Pred. No. 1.4e-159; Indels 103; Gaps 1;

Matches 364; Conservative 1; Mismatches 1;

QY 1 MAADQGIFTNSVTLSPVSGSRTGTLPRFRFRSCDRCHAQIKCTGNKVTGRAPCQRC 60

DB 1 MAADQGIFTNSVTLSPVSGSRTGTLPRFRFRSCDRCHAQIKCTGNKVTGRAPCQRC 60

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: February 15, 2005, 01:55:19 ; Search time 40 Seconds  
(without alignments)  
1128.142 Million cell updates/sec

Title: US-09-974-760B-91  
Perfect score: 2456  
Sequence: 1 MAADQGIFTNSVLTSPVEGS.....FSLARKKHGMLRLNNIPP 469

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 95%  
Listing first 45 summaries

Database : PIR 79: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154.5	6.3	919	2 S45889	probable regulator
2	136	5.5	604	2 JC7252	transcription acti
3	135	5.5	433	2 S69999	sterigmatocystin s
4	132	5.4	1170	2 S30010	probable finger pr
5	131.5	5.4	480	2 T38582	hypothetical funga
6	124.5	5.1	768	2 T37601	probable transcript
7	123.5	5.0	607	2 S63395	probable membrane
8	121.5	4.9	964	2 S48404	probable membrane
9	121	4.9	470	2 S58826	hypothetical prote
10	119	4.8	648	2 T20144	hypothetical prote
11	118.5	4.8	857	2 T40824	probable transcrip
12	117	4.8	473	1 RGVYV3	regulatory protein
13	117	4.8	767	2 T41344	probable zinc-fing
14	116.5	4.7	1445	2 T14913	CAT8 protein - yea
15	115	4.7	6420	2 T30283	polyketide synthas
16	113.5	4.6	650	2 JC7937	transcription acti
17	110.5	4.5	827	2 T39608	zinc finger transc
18	110	4.5	909	2 T06635	hypothetical prote
19	109	4.4	1349	2 T13031	nucleoporin - fru
20	109	4.4	1365	2 T13991	nucleoporin 154 -
21	106.5	4.3	775	2 S41962	thiamin repressibl
22	106.5	4.3	867	2 T41308	hypothetical zinc-
23	106.5	4.3	5327	2 T13564	microtubule-associ
24	106	4.3	743	2 T40521	hypothetical prote
25	106	4.3	1433	2 S54587	CAT8 protein - yea
26	105	4.3	997	2 Ae0776	230k bullous pemph
27	104.5	4.3	1171	2 T00380	XIAA0637 protein -
28	104	4.2	1280	2 T00365	hypothetical prote
29	104	4.2	1587	2 G86467	hypothetical prote

30	103	4.2	704	2 T50303	hypothetical prote
31	103	4.2	765	2 A36332	regulatory protein
32	102.5	4.2	758	2 S46625	finger protein YJL
33	102.5	4.2	2232	2 T34434	hypothetical prote
34	102	4.2	468	2 S46179	regulatory protein p
35	101.5	4.1	1403	2 S24548	homeotic protein p
36	101	4.1	1176	2 T49482	hypothetical prote
37	100.5	4.1	1068	2 S64015	pleiotropic drug r
38	100	4.1	736	2 T06757	hypothetical prote
39	99.5	4.1	547	2 T39478	zinc-finger protei
40	99	4.0	913	2 S61580	probable membrane
41	99	4.0	1048	2 H87721	protein ZC123.2 [i
42	99	4.0	1065	2 S19482	hypothetical prote
43	99	4.0	1425	2 E89303	protein C47E8.8 [i
44	99	4.0	1675	2 T31473	hypothetical prote
45	98.5	4.0	1095	2 T00329	hypothetical prote

## ALIGNMENTS

### RESULT 1

S45889  
probable regulatory protein YBR033w - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein YBR0318  
C;Species: Saccharomyces cerevisiae  
C;Date: 28-Aug-1994 #sequence\_revision 09-Sep-1994 #text\_change 16-Aug-2004  
C;Accession: S45889; S46562  
R;Grivell, L.A.; de Haan, M.; Maat, M.J.; Smits, P.H.M.  
submitted to the Protein Sequence Database, August 1994  
A;Reference number: S45875  
A;Accession: S45889  
A;Molecule type: DNA  
A;Residues: 1-919 <GRI>  
A;Cross-references: UNIPROT:P38073; EMBL:X76078; NID:235902; PIDN:CAA84975.1; PID:953  
A;Experimental source: strain S288C  
R;Smits, P.H.M.; de Haan, M.; Maat, C.; Grivell, L.A.  
Yeast 10(Suppl.A), S75-S80, 1994  
A;Title: The complete sequence of a 33 kb fragment on the right arm of chromosome II f  
ly identified genes and a homologue of the SCOI gene.  
A;Reference number: S46551; MUID:94378725; PMID:8091864  
A;Accession: S46562  
A;Status: nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA  
A;Residues: 1-919 <SMI>  
A;Cross-references: EMBL:X76078; NID:9498748; PIDN:CAA53688.1; PID:9498760  
A;Experimental source: strain S288C  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993  
C;Genetics:  
A;Cross-references: SGB:S0000237  
A;Map position: 2R  
A;Note: YBR033w  
C;Superfamily: GAL4 zinc binuclear cluster homology  
C;Keywords: DNA binding; transmembrane protein; zinc finger  
F;51-90/Domain: GAL4 zinc binuclear cluster homology <GAL4>  
F;56-85/Region: zinc finger  
F;408-424/Domain: transmembrane #status predicted <TM1>  
F;482-504/Domain: transmembrane #status predicted <TM2>  
F;547-563/Domain: transmembrane #status predicted <TM3>  
F;703-720/Domain: transmembrane #status predicted <TM4>  
F;790-806/Domain: transmembrane #status predicted <TM5>

Query Match 6.3%; Score 154.5; DB 2; Length 919;  
Best Local Similarity 23.1%; Pred. No. 0.0012;  
Matches 82; Conservative 52; Mismatches 142; Indels 79; Gaps 14;

QY	2	AADQGIFTNSVLTSPVEGSRTGTLPRRAFRSCDRCHAQIKCTGNKEVTGRAPQRCQ	61
DB	23	SASMGVSVSRNCERCEKTKRKQKASHACQCRKRIKCRFDKH-TG--VQGCGL	79
QY	62	QAGLRC-----VYSERCFPKLQSRADLVLSADPDP-----CL	95
DB	80	EVGEKQFIRVPLKRGPAKKGVSWSIEKFSFSDNDPLQYRPRTHSYPMNSGNNYLPFLAR	139

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 15, 2005, 01:47:14 ; Search time 173 Seconds  
(without alignments)

1388.238 Million cell updates/sec

Title: US-09-974-760B-91

Perfect score: 2456

Sequence: 1 MAADQGIPTSVTLSPVSGS.....FSLARKHKGWLRDLNIPP 469

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612377

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 95%

Listing first 45 summaries

Database :

UniProt\_03.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	830	33.8	459	Q8J0F2	Q8J0F2 penicillium
2	161.5	6.6	1010	Q7S6Q4	Q7S6Q4 neurospora
3	158	6.4	478	Q6BPX4	Q6BPX4 debaryomyce
4	154.5	6.3	919	Y803_YEAST	P38073 saccharomyc
5	148.5	6.0	500	Q6Q877	Q6Q877 leptosphaer
6	145.5	5.9	1082	Q6CX50	Q6CX50 kluyveromyc
7	145	5.9	603	Q7SEN4	Q7SEN4 neurospora
8	142.5	5.8	520	Q8X0T1	Q8X0T1 neurospora
9	138.5	5.6	554	Q7SAV4	Q7SAV4 neurospora
10	137	5.6	1287	Q6FLP2	Q6FLP2 candida gla
11	136	5.5	604	Q60201	Q60201 aspergillus
12	135.5	5.5	442	Q8TGB1	Q8TGB1 aspergillus
13	135	5.4	433	1 AFLR EMENI	P52957 emericella
14	133.5	5.4	922	Q6EX18	Q6EX18 candida gla
15	133.5	5.4	1254	Q6FJW6	Q6FJW6 candida gla
16	133	5.4	470	Q6UVH4	Q6UVH4 saccharomyc
17	132	5.4	1170	1 YK08_YEAST	P28662 saccharomyc
18	131.5	5.4	480	Q94392	Q94392 schizosacch
19	130.5	5.3	662	Q9Y728	Q9Y728 emericella
20	130.5	5.3	785	Q672V8	Q672V8 aspergillus
21	130	5.3	436	Q61WF3	Q61WF3 aspergillus
22	129.5	5.3	579	Q9Y8A1	Q9Y8A1 aspergillus
23	128.5	5.2	442	Q8TG77	Q8TG77 aspergillus
24	128.5	5.2	442	Q8TG79	Q8TG79 aspergillus
25	128.5	5.2	442	Q8TG84	Q8TG84 aspergillus
26	127	5.2	832	Q6FUL5	Q6FUL5 candida gla
27	126.5	5.2	341	1 ACE2_TRIRE	Q96WN6 trichoderma
28	126.5	5.2	442	Q8TFC7	Q8TFC7 aspergillus
29	126.5	5.2	442	Q8TG78	Q8TG78 aspergillus
30	126.5	5.2	442	Q8TG86	Q8TG86 aspergillus
31	126.5	5.2	1353	2 Q7S6K8	Q7S6K8 neurospora

32	125.5	5.1	442	2	Q8TGS2	Q8TGS2 aspergillus
33	124.5	5.1	442	2	Q8TGS3	Q8TGS3 aspergillus
34	124.5	5.1	442	2	Q8TGS7	Q8TGS7 aspergillus
35	124	5.0	984	2	Q94167	Q94167 glomerella
36	123.5	5.0	437	1	AFLR_ASPFL	P41765 aspergillus
37	123.5	5.0	444	2	Q8TG71	Q8TG71 aspergillus
38	123.5	5.0	607	1	YN92_YEAST	P53749 saccharomyc
39	123	5.0	444	2	Q8TG72	Q8TG72 aspergillus
40	122	5.0	441	2	Q8TG80	Q8TG80 aspergillus
41	122	5.0	444	2	Q8J2V0	Q8J2V0 aspergillus
42	122	5.0	444	2	Q874B8	Q874B8 aspergillus
43	122	5.0	821	1	ALCR_EMENI	P21228 emericella
44	121.5	4.9	964	1	YINO_YEAST	P40467 saccharomyc
45	121	4.9	444	2	Q8TFC9	Q8TFC9 aspergillus

## ALIGNMENTS

RESULT 1  
Q8J0F2 PRELIMINARY; PRT; 459 AA.  
AC Q8J0F2;  
DT 01-MAR-2003 (TREMBLrel. 23, Created)  
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Transcription factor.  
DE Name=mlcR;  
GN Penicillium citrinum.  
OS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
OC Eukaryotes; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.  
OX NCBI\_TaxID=5077;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22162865; PubMed=12172803; DOI=10.1007/s00438-002-0697-y;  
RA Abe Y., Suzuki T., Ono C., Iwamoto K., Hosobuchi M., Yoshikawa H.;  
RT "Molecular cloning and characterization of an ML-236B (compactin) biosynthetic gene cluster in Penicillium citrinum."  
RL Mol. Genet. Genomics 268:130-137(2002).  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- SIMILARITY: Contains 1 Zn(2)-Cys(6) fungal-type binuclear cluster domain.  
CC EMBL: AB072893; BAC20569.1; -.  
DR GO: GO:0005634; C:nucleus; IEA.  
DR GO: GO:0003700; F:transcription factor activity; IEA.  
DR GO: GO:0008270; F:zinc ion binding; IEA.  
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
DR InterPro: IPR001138; Fungi\_Trcsrp\_N.  
DR Pfam: PF00172; Zn\_clus\_1.  
DR PRINTS: PR00054; FUNGALZNCYS.  
DR SMART: SM00066; GAL4.1.  
DR PROSITE: PS00463; ZN2\_Cy6\_FUNGAL\_1; UNKNOWN\_1.  
DR PROSITE: PS00448; ZN2\_Cy6\_FUNGAL\_2; 1.  
DR DNA-binding; Metal-binding; Nuclear protein; Transcription;  
KW Transcription regulation; Zinc.  
SQ SEQUENCE 459 AA; 50356 MW; 4D2D1B3307EF46C1 CRC64;

Query Match 33.8%; Score 830; DB 2; Length 459;

Best Local Similarity 42.2%; Pred. No. 3.3e-54;

Matches 196; Conservative 62; Mismatches 153; Indels 54; Gaps 9;

QY 26 LRRAPRRSCDRCHQAQIKCTGKNTGVPAPQRCQQAGLCVYSERCPRKLRQRAA- 84

DB 12 LRRAPRRSCDRCHQAQIKCTGKNTGVPAPQRCQQAGLCVYSERLPKRLHKEAAG 71

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 20:20:52 ; Search time 5100 seconds

(without alignment)  
4455.980 Million cell updates/sec

Title: US-09-974-760B-91

Perfect score: 2456

Sequence: 1 MAADQGIPTNSVTLSPVEGS.....FSLARKKHGMRDLNIPP 469

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 95%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US09974760/runat\_14022005\_160342\_6751/app\_query.fasta\_1.647  
-DB=genembi -QFMT=fastap -SUFFIX=95pct.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bites -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=95 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09974760 @CGN 1 1 5600 @runat\_14022005\_160342\_6751 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb\_ba.\*  
2: gb\_hcg.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	830	33.8	1380	6	BD178141
C 2	811	33.0	34203	6	BD178102
3	811	33.0	34203	6	BD178103
C 4	811	33.0	34203	6	BD013765

5	811	33.0	34203	6	BD013766
C 6	811	33.0	34203	6	BD093553
7	811	33.0	34203	6	BD093554
C 8	811	33.0	38231	8	AB072893
C 9	292.5	11.9	562	6	BD178135
C 10	292.5	11.9	562	6	BD013798
C 11	292.5	11.9	562	6	BD093586
C 12	266.5	10.9	541	6	BD178123
C 13	266.5	10.9	541	6	BD013786
C 14	266.5	10.9	541	6	BD093574
15	163	6.6	2757	6	AX684918
16	163	6.6	3114	8	SCYBR033W
C 17	163	6.6	3117	8	SCGAL1
C 18	161	6.6	68683	8	AY553235
19	158	6.4	110000	8	CR382137_07
20	157.5	6.4	1272	6	AX505237
21	157.5	6.4	1272	6	AX684982
22	153.5	6.2	1440	6	AX684768
23	153.5	6.2	11578	8	SPAC2H10
24	146	5.9	241050	2	AC119506
C 25	146	5.9	244250	2	AC095307
C 26	145.5	5.9	110000	8	CR382121_07
C 27	145.5	5.9	35745	3	LMFL5174
28	145	5.9	3667	8	AK110197
29	143.5	5.8	3667	8	AK110197
C 30	143.5	5.8	261960	2	AC120634
C 31	140.5	5.7	1812	6	AX684724
32	140	5.7	2045	8	AF441422
33	140	5.7	2051	8	AF441414
34	140	5.7	110000	2	AP006487_2
35	140	5.7	301950	1	AP006570
36	139.5	5.7	110000	8	CR380958_02
C 37	139.5	5.7	245684	2	AC098110
C 38	139.5	5.7	255473	2	AC121416
39	139	5.7	3980	6	AR309623
40	139	5.7	3980	6	AR309624
41	139	5.7	3986	8	AOAJ5258
42	139	5.7	8813	8	AB012945
C 43	139	5.7	35244	8	AB021876
C 44	138.5	5.6	1934	3	LTALED
C 45	138.5	5.6	301675	1	AP005027

#### ALIGNMENTS

RESULT 1	BD178141	1380 bp	DNA	linear	PAT 16-APR-2003
LOCUS	On the structural gene of gene cluster.				
DEFINITION	BD178141				
ACCESSION	BD178141				
VERSION	BD178141.1	GI:30015405			
KEYWORDS	JP 2002315579-A/40.				
SOURCE	Penicillium citrinum				
ORGANISM	Penicillium citrinum				
REFERENCE	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
AUTHORS	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.				
TITLE	1 (bases 1 to 1380)				
JOURNAL	Abe, I., Ono, C. and Yoshikawa, H.				
COMMENT	On the structural gene of gene cluster				
	Patent: JP 2002315579-A 40 29-OCT-2002;				
	SANKYO CO LTD				
	OS Penicillium citrinum				
	PN JP 2002315579-A/40				
	PD 29-OCT-2002				
	PF 18-APR-2001				
	PI YUKI ABE, CHIHO ONO, HIROJI YOSHIKAWA				
	PC C12N1/09, C12N1/15, C12N1/21, C12P17/06, C12N1/15, C12R1/80, PC				
	(C12P17/06, C12R1/80) C12N15/00				
	CC On the structural gene of gene cluster				
	PH Key				
	FT Location/Qualifiers				
	FT CDS				
	1..1380				

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 19:56:52 ; Search time 643 Seconds

(without alignment)  
4317.821 Million cell updates/sec

Title: US-09-974-760B-91

Perfect score: 2456

Sequence: 1 MAADQGIPTSVTLSPVEGS.....FSLARKKHGMLRLNIPP 469

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780382

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 95%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xip  
-Q=/cgn2\_1/USFIO\_epool\_p/US09974760/runat\_14022005\_160342\_6744/app\_query.fasta\_1.647  
-DB=N Geneseq 16Dec04 -Qfmt=fastap -SUFFIX=95pct.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=95 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US09974760 @CGN 1 1 708 @runat\_14022005\_160342\_6744 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 16Dec04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2002bs:\*  
8: Geneseqn2003as:\*  
9: Geneseqn2003bs:\*  
10: Geneseqn2003cs:\*  
11: Geneseqn2003ds:\*  
12: Geneseqn2004as:\*  
13: Geneseqn2004bs:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	830	33.8	1377	10 ADE81190	Ade81190 MLCr codi
2	830	33.8	1380	6 AAF74547	Aaf74547 Penicilli
c 3	811	33.0	34203	4 AAF74546	Aaf74546 Penicilli
c 4	811	33.0	34203	6 AAF74547	Aaf74547 Penicilli
5	811	33.0	34263	4 AAF74547	Aaf74547 Penicilli

c 6	811	33.0	72149	10 ADE81173	Ade81173 ML-236B s
c 7	292.5	11.9	562	4 AAF74579	Aaf74579 ML-236B b
c 8	292.5	11.9	562	6 AAF74587	Aaf74587 ML-236B b
c 9	266.5	10.9	541	6 AAF74587	Aaf74587 ML-236B b
10	266.5	10.9	541	6 AAF74587	Aaf74587 ML-236B b
11	163	6.6	2757	6 AAF74587	Aaf74587 ML-236B b
12	163	6.6	2757	6 AAF74587	Aaf74587 ML-236B b
13	157.5	6.4	1272	6 AAF74587	Aaf74587 ML-236B b
14	157.5	6.4	1272	6 AAF74587	Aaf74587 ML-236B b
15	153.5	6.2	1440	6 AAF74587	Aaf74587 ML-236B b
16	140.5	5.7	1812	6 AAF74587	Aaf74587 ML-236B b
17	139	5.7	3980	2 AAF74587	Aaf74587 ML-236B b
18	135	5.5	1299	6 AAF74587	Aaf74587 ML-236B b
19	132.5	5.4	2100	2 AAF74587	Aaf74587 ML-236B b
20	132	5.4	1311	6 AAF74587	Aaf74587 ML-236B b
21	132	5.4	3510	6 AAF74587	Aaf74587 ML-236B b
22	132	5.4	3513	13 AAF74587	Aaf74587 ML-236B b
23	130.5	5.3	1986	6 AAF74587	Aaf74587 ML-236B b
24	130.5	5.3	2384	6 AAF74587	Aaf74587 ML-236B b
25	129.5	5.3	1737	6 AAF74587	Aaf74587 ML-236B b
26	129	5.3	1152	6 AAF74587	Aaf74587 ML-236B b
27	129	5.3	1332	6 AAF74587	Aaf74587 ML-236B b
28	128.5	5.2	1373	2 AAF74587	Aaf74587 ML-236B b
c 29	127.5	5.2	2170	6 AAF74587	Aaf74587 ML-236B b
c 30	127.5	5.2	5378	13 AAF74587	Aaf74587 ML-236B b
c 31	127.5	5.2	5382	10 AAF74587	Aaf74587 ML-236B b
32	126.5	5.2	3176	10 AAF74587	Aaf74587 ML-236B b
33	126.5	5.2	3199	13 AAF74587	Aaf74587 ML-236B b
34	126.5	5.2	3312	13 AAF74587	Aaf74587 ML-236B b
35	126.5	5.2	3996	12 AAF74587	Aaf74587 ML-236B b
36	126.5	5.2	5154	13 AAF74587	Aaf74587 ML-236B b
37	126.5	5.2	7819	10 AAF74587	Aaf74587 ML-236B b
38	126	5.1	2466	6 AAF74587	Aaf74587 ML-236B b
39	125.5	5.1	2217	6 AAF74587	Aaf74587 ML-236B b
40	124.5	5.1	2208	6 AAF74587	Aaf74587 ML-236B b
41	124	5.0	2952	6 AAF74587	Aaf74587 ML-236B b
42	123.5	5.0	1821	6 AAF74587	Aaf74587 ML-236B b
43	123	5.0	1596	10 AAF74587	Aaf74587 ML-236B b
44	123	5.0	1695	6 AAF74587	Aaf74587 ML-236B b
45	122.5	5.0	2646	6 AAF74587	Aaf74587 ML-236B b

## ALIGNMENTS

RESULT 1	
ID ADE81190	standard; DNA; 1377 BP.
XX	
AC ADE81190;	
XX	
DT 29-JAN-2004	(first entry)
XX	
DE MLCr coding sequence, SEQ ID 18.	
XX	
KW ML-236B; HMG-CoA reducing enzyme; ds; MLCr.	
XX	
OS Penicillium citrinum.	
XX	
FT Key	Location/Qualifiers
FT CDS	1..1377
FT	/*tag= a
FT	/product= "MLCr"
XX	
DN JP2003116567-A.	
XX	
PD 22-APR-2003.	
XX	
PF 15-OCT-2001; 2001JP-00316578.	
XX	
PR 15-OCT-2001; 2001JP-00316578.	
XX	
PA (SANY ) SANKYO CO LTD.	
XX	



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 21:08:47 ; Search time 234 Seconds  
(without alignment)  
3279.546 Million cell updates/sec

Title: US-09-974-760B-91

Perfect score: 2456

Sequence: 1 MAADQGIPTSVTLSPVSGS.....FSLARKKHGMRDLNNIPP 469

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop. 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405566

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 95%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlip  
-Q=/cgn2\_1/USPTO\_spool\_p/US09974760/runat\_14022005\_160343\_6778/app\_query.fasta\_1.647  
-DB=Issued Patents NA -Qfmt=fastap -SUFFIX=95pct.rn1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=95 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09974760 @CIGN 1 1 105 @runat\_14022005\_160343\_6778 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Databases :

- 1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*
- 2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*
- 3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*
- 4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*
- 5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*
- 6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	157.5	6.4	1272	US-10-029-180-103	Sequence 103, App
2	139	5.7	3980	US-09-197-814-1	Sequence 1, Appli
3	139	5.7	3980	US-09-197-814-2	Sequence 2, Appli
4	139	5.7	3980	US-09-920-581-1	Sequence 1, Appli
5	139	5.7	3980	US-09-920-581-2	Sequence 2, Appli
6	129.5	5.3	49225	US-09-902-540-1269	Sequence 1269, Ap
7	127	5.2	2844	US-09-669-980A-5	Sequence 5, Appli
8	126.5	5.2	3061	US-09-147-119-6	Sequence 6, Appli
9	126.5	5.2	3145	US-09-949-016-2624	Sequence 2624, Ap
10	121.5	4.9	2970	US-09-248-796A-5006	Sequence 5006, Ap
11	121	4.9	110585	US-09-949-016-13427	Sequence 13427, A
12	119.5	4.9	4403765	US-09-103-840A-2	Sequence 2, Appli

13	119.5	4.9	4411529	3	US-09-103-840A-1	Sequence 1, Appli
14	119	4.8	3047	4	US-09-016-434-1465	Sequence 1465, Ap
15	118	4.8	927	4	US-10-029-180-97	Sequence 97, Appl
16	118	4.8	1317	4	US-09-248-796A-4861	Sequence 4861, Ap
17	118	4.8	4403765	3	US-09-103-840A-1	Sequence 1, Appli
18	118	4.8	4411529	3	US-09-103-840A-2	Sequence 2, Appli
19	114.5	4.7	2961	4	US-09-248-796A-4985	Sequence 4985, Ap
20	113.5	4.6	3226	3	US-08-870-126-10	Sequence 10, Appl
21	113.5	4.6	3226	3	US-09-445-247-10	Sequence 10, Appl
22	112.5	4.6	2855	4	US-09-248-796A-6094	Sequence 6094, Ap
23	112.5	4.6	4935	4	US-09-949-016-1019	Sequence 1019, Ap
24	112.5	4.6	4935	4	US-09-949-016-5054	Sequence 5054, Ap
25	112.5	4.6	41062	4	US-09-949-016-12761	Sequence 12761, A
26	112.5	4.6	41062	4	US-09-949-016-16796	Sequence 16796, A
27	112	4.6	109690	4	US-09-949-016-13525	Sequence 13525, A
28	111.5	4.5	44377	2	US-08-804-227C-7	Sequence 7, Appli
29	111.5	4.5	44377	2	US-08-804-198-1	Sequence 1, Appli
30	110.5	4.5	792	4	US-09-252-991A-15292	Sequence 15292, A
31	110	4.5	2239	4	US-09-774-528-49	Sequence 49, Appl
32	109.5	4.5	2871	4	US-09-252-991A-11431	Sequence 11431, A
33	109.5	4.5	9880	4	US-09-902-540-936	Sequence 936, App
34	109	4.4	1548	4	US-09-252-991A-5880	Sequence 5880, Ap
35	109	4.4	1605	4	US-09-248-796A-5465	Sequence 5465, Ap
36	109	4.4	1735	3	US-09-163-444-3	Sequence 3, Appli
37	108.5	4.4	942	4	US-09-248-796A-4573	Sequence 4573, Ap
38	108.5	4.4	2052	4	US-09-252-991A-12103	Sequence 12103, A
39	108.5	4.4	5895	4	US-09-902-540-7920	Sequence 792, App
40	108.5	4.4	33578	4	US-09-949-016-15670	Sequence 15670, A
41	108	4.4	2508	4	US-09-252-991A-15550	Sequence 15550, A
42	107.5	4.4	2847	4	US-09-902-540-7588	Sequence 7588, Ap
43	107.5	4.4	4978	4	US-09-902-540-775	Sequence 775, App
44	107.5	4.4	7704	4	US-09-902-540-743	Sequence 743, App
45	107.5	4.4	10322	4	US-09-902-540-989	Sequence 989, App

## ALIGNMENTS

RESULT 1

US-10-029-180-103  
; Sequence 103, Application US/10029180  
; Patent No. 6806082  
; GENERAL INFORMATION:  
; APPLICANT: Cali, Brian M.  
; APPLICANT: Holtzman, Doug  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: Milna, G. Todd  
; APPLICANT: Sherman, Amir  
; APPLICANT: Silva, Jeffrey C.  
; APPLICANT: Trueheart, Josh  
; APPLICANT: Zhang, Lixin  
; TITLE OF INVENTION: No. 6806082el Regulators of Fungal Gene Expression  
; FILE REFERENCE: MIC-004  
; CURRENT APPLICATION NUMBER: US/10/029,180  
; CURRENT FILING DATE: 2001-12-22  
; PRIOR APPLICATION NUMBER: US 60/257,431  
; PRIOR FILING DATE: 2000-12-22  
; NUMBER OF SEQ ID NOS: 138  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 103  
; LENGTH: 1272  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: fungal gene  
US-10-029-180-103

Alignment Scores:  
Pred. No.: 2,53e-06 Length: 1272  
Score: 157.50 Matches: 95  
Percent Similarity: 36.65% Conservative: 67  
Best Local Similarity: 21.49% Mismatches: 170  
Query Match: 6.41% Indels: 111  
DB: 4 Gaps: 19

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: February 16, 2005, 22:54:23 ; Search time 673 Seconds  
(without alignments)  
4111.910 Million cell updates/sec

Title: US-09-974-760B-91  
Perfect score: 2456  
Sequence: 1 MAADQGGFTSVTLSPVSGS.....FSLARKKHGMLRLNIPP 469

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 5378673 seqs, 2950229984 residues

Total number of hits satisfying chosen parameters: 10757281

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 95%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US09974760/runat\_14022005\_160344\_6868/app\_query.fasta\_1.647  
-DB=Published Applications NA -QFMT=fastap -SUFFIX=95pct.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blom62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=95  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -SCOR=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09974760 @CIGN 1 1 723 @runat 14022005 160344 6868  
-NCPU=6 -ICPU=3 -NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-XGAPOP=6 -XGAPEXT=7 -XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq.\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq.\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq.\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq.\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq.\*  
15: /cgn2\_6/ptodata/2/pubpna/US10C\_PUBCOMB.seq.\*  
16: /cgn2\_6/ptodata/2/pubpna/US10D\_PUBCOMB.seq.\*  
17: /cgn2\_6/ptodata/2/pubpna/US10E\_PUBCOMB.seq.\*  
18: /cgn2\_6/ptodata/2/pubpna/US10F\_PUBCOMB.seq.\*  
19: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq.\*  
20: /cgn2\_6/ptodata/2/pubpna/US11\_NEW\_PUB.seq.\*  
21: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq.\*  
22: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1851.5	75.4	1101	18	US-10-402-056-114	Sequence 114, Appl
2	830	33.8	1380	10	US-09-836-705-41	Sequence 41, Appl
3	811	33.0	34203	10	US-09-836-705-1	Sequence 1, Appl
4	811	33.0	34203	10	US-09-836-705-2	Sequence 2, Appl
5	292.5	11.9	562	10	US-09-836-705-34	Sequence 34, Appl
6	266.5	10.9	541	10	US-09-836-705-22	Sequence 22, Appl
7	163	6.6	2757	17	US-10-149-310-337	Sequence 237, Appl
8	163	6.6	2760	17	US-10-149-310-337	Sequence 237, Appl
9	157.5	6.4	1272	13	US-10-029-180-103	Sequence 103, Appl
10	157.5	6.4	1272	17	US-10-149-310-301	Sequence 301, Appl
11	153.5	6.2	1440	17	US-10-149-310-87	Sequence 87, Appl
12	140.5	5.7	1812	17	US-10-149-310-45	Sequence 45, Appl
13	139	5.7	3980	9	US-09-920-581-1	Sequence 1, Appl
14	139	5.7	3980	9	US-09-920-581-2	Sequence 2, Appl
15	139	5.7	3980	15	US-10-371-421-1	Sequence 1, Appl
16	139	5.7	3980	15	US-10-371-421-2	Sequence 2, Appl
17	138.5	5.6	9025608	15	US-10-156-761-1	Sequence 1, Appl
18	136	5.5	1557	18	US-10-437-963-92795	Sequence 92795, A
19	135	5.5	1299	17	US-10-149-310-39	Sequence 29, Appl
20	135	5.5	1302	9	US-09-801-368-37	Sequence 37, Appl
21	133.5	5.4	1551	15	US-10-156-761-1622	Sequence 1622, Ap
22	132	5.4	1311	17	US-10-149-310-27	Sequence 27, Appl
23	132	5.4	3510	17	US-10-149-310-197	Sequence 197, Appl
24	132	5.4	3513	17	US-10-369-493-45946	Sequence 45946, A
25	130.5	5.3	1866	17	US-10-149-310-43	Sequence 43, Appl
26	130.5	5.3	1989	9	US-09-801-368-41	Sequence 41, Appl
27	130.5	5.3	2384	18	US-10-468-250A-78	Sequence 78, Appl
28	129.5	5.3	1737	17	US-10-149-310-31	Sequence 31, Appl
29	129	5.3	1152	17	US-10-149-310-31	Sequence 31, Appl
30	129	5.3	1332	17	US-10-149-310-33	Sequence 33, Appl
31	127.5	5.2	2170	17	US-10-264-049-850	Sequence 850, Appl
32	127	5.2	894	19	US-10-943-742-5	Sequence 5, Appl
33	126.5	5.2	284	18	US-10-437-963-21580	Sequence 21580, A
34	126	5.1	2466	15	US-10-468-250A-74	Sequence 74, Appl
35	126	5.1	9025608	15	US-10-156-761-1	Sequence 1, Appl
36	125.5	5.1	2217	17	US-10-149-310-213	Sequence 213, Appl
37	125.5	5.1	2220	9	US-09-801-368-395	Sequence 395, Appl
38	124.5	5.1	2208	17	US-10-149-310-91	Sequence 91, Appl
39	124	5.0	2952	17	US-10-149-310-117	Sequence 117, Appl
40	123.5	5.0	1821	17	US-10-149-310-187	Sequence 187, Appl
41	123	5.0	1695	17	US-10-149-310-121	Sequence 121, Appl
42	122.5	5.0	2646	17	US-10-149-310-121	Sequence 121, Appl
43	122.5	5.0	17493	19	US-10-211-028-10	Sequence 10, Appl
44	122.5	5.0	90597	19	US-10-211-028-1	Sequence 1, Appl
45	122	5.0	2463	17	US-10-149-310-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1  
US-10-402-056-114  
; Sequence 114, Application US/10402056  
; Publication No. US20040191877A1  
; GENERAL INFORMATION:  
; APPLICANT: Roberts, Shannon  
; APPLICANT: Sherman, Amir  
; APPLICANT: Trueheart, Joshua  
; APPLICANT: Milne, G. Todd  
; TITLE OF INVENTION: LOVE VARIANT REGULATOR MOLECULES  
; FILE REFERENCE: 14184-029001  
; CURRENT APPLICATION NUMBER: US/10402,056  
; CURRENT FILING DATE: 2003-03-28  
; PRIOR APPLICATION NUMBER: PCT/US02/32248  
; PRIOR FILING DATE: 2002-10-09  
; PRIOR APPLICATION NUMBER: US 09/974,760  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: US 60/328,339  
; PRIOR FILING DATE: 2001-10-09  
; NUMBER OF SEQ ID NOS: 118

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: February 16, 2005, 21:02:37 ; Search time 3602 Seconds  
(without alignments)  
4956.175 Million cell updates/sec

Title: US-09-974-760B-91  
Perfect score: 2456  
Sequence: 1 MAADQIGFTSVTLSPVEGS.....FSLARKKHGMRLDNLNIPP 469

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 95%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame+p2n.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_spool\_p/US099974760/runat\_14022005\_160343\_6765/app\_query.fasta\_1.647  
-DB=EST -QPMT=fastap -SUFFIX=95pct.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR SCORE=pct -THR MAX=95 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09974760@cgn\_1\_1\_5180@runat\_14022005\_160343\_6765 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -DSFBLCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_hc:\*  
4: gb\_est3:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_gss1:\*  
9: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	145.5	5.9	791	CF682666	CF682666 CCACQ27TR
2	143.5	5.8	705	BQ110832	BQ110832 VD0205D08
3	136	5.5	925	C0008256	C0008256 EST796591
4	131	5.3	773	CF706713	CF706713 CCAPSS59TR
5	130.5	5.3	570	AQ397018	AQ397018 mgxb0021J
6	129.5	5.3	1503	BM918679	BM918679 AGENCOURT
7	129	5.3	2472	CL961584	CL961584 OaIFCC006
8	127.5	5.2	731	CF868238	CF868238 trico14xd
9	127.5	5.2	786	CB898308	CB898308 trico14xd

10	127.5	5.2	857	7	C0021754	C0021754
11	127.5	5.2	2066	9	AY399870	AY399870 Homo sapi
12	127.5	5.2	2274	3	CR590353	CR590353 full-leng
13	126.5	5.2	740	7	CF886493	CF886493 trico10xh
14	126.5	5.2	812	6	CB902489	CB902489 trico10xh
15	126.5	5.2	1904	3	BC050362	BC050362 Homo sapi
16	126.5	5.2	5176	3	CR749286	CR749286 Homo sapi
17	126	5.1	584	7	C0134902	C0134902 EST829573
18	125	5.1	623	7	C0022915	C0022915 EST819008
19	125	5.1	1094	6	CD458335	CD458335 F808_09e0
20	123.5	5.0	1066	6	CD501980	CD501980 CDA5J-H11
21	123	5.0	2786	3	AK082888	AK082888 Mus muscu
22	122	5.0	925	7	C0021563	C0021563 EST817656
23	120.5	4.9	750	7	CF877407	CF877407 trico075xp
24	120.5	4.9	765	7	CF688031	CF688031 CCAI871TR
25	120.5	4.9	812	6	CB905758	CB905758 trico075xp
26	120.5	4.9	1272	5	BQ648956	BQ648956 AGENCOURT
27	120	4.9	706	5	BQ063507	BQ063507 Fgr_3_D08
28	119.5	4.9	1530	4	BM543030	BM543030 AGENCOURT
29	119	4.8	683	7	CF693020	CF693020 CCAAH76TR
30	119	4.8	695	7	CF870415	CF870415 trico022xn
31	119	4.8	766	6	CB900596	CB900596 trico022xn
32	118.5	4.8	858	7	CN809042	CN809042 EST0388 M
33	118	4.8	759	7	CF869789	CF869789 trico020xh
34	118	4.8	815	6	CB899935	CB899935 trico020xh
35	118	4.8	974	9	CNS03ADZ	AL235088 Tetraodon
36	117.5	4.8	1632	9	CL945327	CL945327 OaIFSB003
37	117.5	4.8	382	9	CL957631	CL957631 OaIFCC036
38	117	4.8	558	5	BQ138028	BQ138028 NF012D11P
39	117	4.8	735	7	CF714957	CF714957 CCADQ07TF
40	117	4.8	818	7	CO141871	CO141871 EST836542
41	117	4.8	866	7	CF711138	CF711138 CCAF326TO
42	117	4.8	947	9	CC668530	CC668530 OGLAV40TH
43	116.5	4.7	789	7	CF687839	CF687839 CCAA796TR
44	116.5	4.7	844	7	CF695274	CF695274 CCADBI3TR
45	116.5	4.7	873	7	CF694730	CF694730 CCAE246TR

ALIGNMENTS

RESULT 1  
CF682666 791 bp mRNA linear EST 16-AUG-2004  
LOCUS CCACQ27TR C.neoformans strain JEC21 Cryptococcus neoformans var.  
neoformans cDNA clone CCACQ27, mRNA sequence.

ACCESSION CF682666  
VERSION CF682666  
KEYWORDS EST.  
SOURCE CF682666.1 GI:41536825  
Cryptococcus neoformans var. neoformans (Filobasidiella neoformans var. neoformans)

ORGANISM  
Cryptococcus neoformans var. neoformans  
Eukaryota; Fungi; Basidiomycota; Hymenomycetes;  
Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae;  
Filobasidiella.

REFERENCE  
1 (bases 1 to 791)  
Loftus, B.  
End sequencing of clones from a Full length enriched, normalized  
JEC21 cDNA library

JOURNAL  
Unpublished (2003)  
COMMENT  
Other\_ESTs: CCACQ27TF  
Contact: Brendan Loftus

TIGR  
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,  
Rockville, MD 20850, USA  
Tel: 301-838-3543  
Fax: 301-838-0208  
Email: crypt@tigr.org

Seq primer: TR.  
Location/Qualifiers  
1..791  
/organism="Cryptococcus neoformans var. neoformans"  
/mol\_type="mRNA"  
/strain="JEC21"

FEATURES  
source  
CF682666